

---

# SYSTEMATICS INFORMATION AS A CENTRAL COMPONENT IN THE NATIONAL BIOLOGICAL INFORMATION INFRASTRUCTURE

---

Anne Frondorf<sup>1</sup> and Gary Waggoner<sup>2</sup>

---

## ABSTRACT

The National Biological Service (NBS) is leading the development of the National Biological Information Infrastructure (NBII), a distributed network of biological data and information available electronically over the Internet. The important role of systematics information in the NBII is discussed by presenting a vision of a "national systematics information infrastructure." Several current initiatives of the NBS that directly contribute to developing such a systematics information infrastructure are described.

---

The mission of the National Biological Service (NBS) is to work with other agencies and organizations to provide the scientific understanding and technologies needed to manage the Nation's (i.e., the United States') biological resources. A fundamental part of this mission is to make data and information on biological resources more accessible for more people to use in making resource management decisions. (For the purposes of this discussion, the term "data" refers to material as originally collected or developed, while "information" refers more broadly to material that has been processed, or integrated with other material, or displayed or presented in certain ways. The objective of the NBS and the National Biological Information Infrastructure is to make *both* data and information more broadly accessible.) A key element of the NBS program is development of a national partnership for sharing biological information: the National Biological Information Infrastructure (NBII). The NBII concept is a significant component of the recommendations made by the National Research Council in its 1993 report entitled "A Biological Survey for the Nation."

The goal of the NBII is to establish a distributed "federation" of biological data and information sources, relying on a network of partners and co-operators to make the data they generate and/or maintain available to others throughout this federation, using the Internet. The basic NBII philosophy is to encourage and facilitate biological data stewardship. Under this philosophy, the management of databases remains with those who are re-

sponsible for the data, using the "information superhighway" to interconnect and disseminate these data to others.

In addition to biological data and information, new types of software tools providing new capabilities will also be available over the NBII network, to help users work at their own local computer with the actual data from several different remote sources, or to work collaboratively with others at remote sites by using the network (i.e., virtual laboratories). The NBII will also point to sources of biological science expertise: people and organizations that users can contact to get assistance in finding and understanding biological data.

The NBII is being implemented in three sequential phases. The NBII Directory is currently pointing users to biological databases and information sources, within NBS and from other agencies or organizations. The NBII Clearinghouse is also now in operation, with several sites serving metadata describing and documenting their data holdings and, in some cases, also the data themselves. (The term "metadata" refers to data which serve to describe other data, including such things as how, when, where, and by whom the data were collected; the subject matter of the data; indications of data quality; and information on how to obtain or access the data. In a library, for example, if the books are considered "data," the information in the card catalog would be the "metadata.") The NBII Distributed System is the ultimate goal of the NBII and is expected to be operational by 1998. In this phase, users will be able to directly access, retrieve, com-

---

<sup>1</sup> National Biological Service, U.S. Department of the Interior, 1849 C St. NW, Washington D.C. 20240, U.S.A.

<sup>2</sup> National Biological Service, U.S. Department of the Interior, Technology Transfer Center, P.O. Box 25387, Building 810, DFC, Denver, Colorado 80225, U.S.A.



bine, and analyze biological data from different sources at different locations (i.e., a virtual national biological database).

#### CURRENT EMPHASES IN NBII DEVELOPMENT

NBS is working simultaneously on three fronts in implementing NBII: (1) making the most significant biological data and information products from NBS research and inventory programs electronically accessible; (2) supporting the efforts of many public and private partners to make their significant biological data and information accessible to others over the NBII; and (3) working cooperatively with other agencies and organizations on the tools and standards needed to provide the infrastructure components of NBII.

Examples of significant biological data and information currently available from NBS through NBII include the North American Breeding Bird Survey, data from the nationwide GAP analysis program, wildlife health bulletins, and data on non-indigenous aquatic species, including the zebra mussel. We continue to work with the NBS science centers, cooperative research units, and other offices to identify NBS biological data and information holdings, and to systematically proceed with cataloging, describing, and electronically serving the most significant of these on NBII.

Partner agencies and organizations, including federal and state agencies, private groups, universities, museums, herbaria, and libraries, have by far the bulk of the biological data and information in the country. Thus, we are working with many partners to make the biological data and information they have more accessible over NBII. Key areas of interest for NBS in developing these partnerships are with other federal agencies, with state resource agencies, and with the systematics and collections management community.

Many of these partners already are making data and information accessible over the Internet, and NBII is linking to these providers, such as the National Wetlands Inventory database of the U.S. Fish and Wildlife Service, the Global Change Master Directory of data and information, and metadata and data in the National Spatial Data Infrastructure. Other agencies and organizations are interested in working with NBS to help automate, describe, and serve their information, some of which may not currently be available in electronic format.

NBS is also working with partners on developing, adapting, and refining the types of software tools, protocols, and standards that are needed to allow users to access, compare, exchange, and use a va-

riety of different data and information sets throughout a widely distributed electronic federation of sources.

NBS has taken a leadership role in developing a draft metadata content standard for use in documenting biological resource data and information in NBII. A metadata content standard specifies what metadata items (fields) to use in describing a data set and how these fields should be formatted and presented. Standardizing the format and presentation of metadata in this way makes it easier for people to compare and contrast among different distributed data sources. The draft standard provides an overall framework for all biological metadata, linking with the Federal Geographic Data Committee standard for spatial data, as well as with the U.S. machine readable cataloging (USMARC) standard for library materials/publications. NBS has contracted with the American Institute for Biological Sciences (AIBS), which assembled a distinguished panel of biological scientists that has provided peer review and recommendations on this draft standard. The final AIBS report has been completed, and NBS will be testing and evaluating this standard and will provide it to other partners who might also be willing to test the standard and provide comments for NBS to use in further revision.

We are also developing and testing new types of software tools and approaches for serving and accessing biological data and information over the Internet. This includes tools for distributed searches, distributed computing, and development of new controlled vocabulary systems to use in searches.

#### IMPORTANCE OF SYSTEMATICS DATA AND INFORMATION IN NBII

Systematics data and information are of fundamental importance in building the NBII. Systematics data and information are necessary to support biological inventory, monitoring, and research. Systematics collections data can be applied in long-term biological or ecological research or to analyze long-term environmental trends. Use of standardized taxonomic nomenclature supports data sharing and comparison between agencies and organizations. Museum collections also provide essential outreach and education services to raise public awareness and understanding of biodiversity.

Systematics data are collected and maintained throughout a widely distributed network of collections, museums, herbaria, botanical gardens, universities, natural history surveys, and other public and private agencies and organizations. In many cases, availability of different systematics data and



information may not be fully known by the people who would be able to use it. Also, much useful data, such as the information on biological specimens housed in museums, are not currently available in an automated format and therefore are not easily accessible. In addition, integration of information from two or more different sources to help answer questions does not always occur because the data from different sources cannot be easily compared or combined.

It is clear that systematics represents a microcosm of all of the issues involved in implementing the NBII. Building on the overall NBII effort, our goal is to help develop and support a "national systematics information infrastructure" in which data, information, and tools are electronically accessible in a distributed network for a broad array of different applications by a variety of users.

#### OBJECTIVES IN DEVELOPING A NATIONAL SYSTEMATICS INFORMATION INFRASTRUCTURE

1. Support the development and use of data and metadata standards, protocols, and tools or technologies to increase the ability to access, understand, share, compare, exchange, and use systematics data and information from many different data and information providers.

2. Support the development of comprehensive surveys and resource directories of the sources, status, and extent of systematics expertise, data, and information.

3. Develop and implement a program that combines small, targeted tactical investments in certain areas where there are opportunities for pilot or demonstration projects, with larger, more strategic investments on major fronts.

4. Actively pursue and develop partnerships in all areas to help meet mutual needs.

#### CURRENT NBII INITIATIVES IN THE SYSTEMATICS INFORMATION ARENA

##### INFRASTRUCTURE DEVELOPMENT

NBS has several activities under way to support development of the tools and standards needed to support a nationally (and internationally) distributed systematics information infrastructure. NBS is an active participant (along with several other federal and state agencies, private organizations, and international organizations) in the development of the Interagency Taxonomic Information System (ITIS). ITIS is the first comprehensive national database to provide quality information on the scientific nomenclature and taxonomy of all U.S. flora

and fauna. ITIS can be accessed on the World Wide Web at <http://www.itis.usda.gov/itis>.

As described above, NBS has developed a proposed metadata content standard for use in documenting non-geospatial biological data and information on NBII. As this proposed standard is refined, tested, and released over the next year, NBS is working to ensure that it includes appropriate elements on systematics/taxonomy and that the proposed standard is beta tested on actual selected systematics data sets to review its utility and applicability. NBS will also work to disseminate information on applicability and use of a metadata standard to the systematics community.

NBS is also working with numerous partners on several fronts to develop a Standardized National Vegetation Classification System. The major effort is in conjunction with two NBS vegetation mapping programs, the National Park Vegetation Mapping program and the National GAP Analysis Program. NBS is working closely with The Nature Conservancy and the network of State Heritage Programs in the development of this draft classification. NBS is also actively working with other federal agencies on the Vegetation Subcommittee of the Federal Geographic Data Committee to develop and adopt national standards for vegetation classification and analysis. Development of these standards will support the coordinated production of uniform statistics on and mapping of the nation's vegetation among all the federal agencies and other cooperators. By sharing in a common reference system for how vegetation cover is classified and described, agencies and organizations that collect and analyze data about vegetation and distribution, trends, etc., can more easily share and exchange their data. And finally, NBS serves on a special panel of the Ecological Society of America charged with assisting in the evaluation and development of a standardized national vegetation classification system.

NBS is also working to foster development, enhancement, and transfer of tools, technologies, and approaches to support the national (and international) systematics information infrastructure. Key areas where tools and approaches are needed include: (1) cost-efficient approaches to large-scale automation of specimen data that are not currently automated; (2) georeferencing of specimen data that are not currently georeferenced; and (3) visualization and imaging tools to allow for electronic "viewing" of specimens.

One example of the work that NBS is supporting in this area is a project to convert vertebrate distribution records of the original Biological Survey in the Smithsonian's National Museum of Natural



History to digital format, with digital geographic coordinate data from a digital gazetteer that links geographic place names to geographic coordinates. The locality database will be available electronically through NBII. Next, records from this biological survey locality system will be matched to localities associated with specimen records in the National Museum's vertebrate specimen databases. This will add geographic coordinates to all matched vertebrate specimen records in the National Museum's database.

Under the NBS State Partnership Program, a competitive program through which NBS supports state-level biodiversity research, inventory, and information transfer activities, NBS is supporting work at the Florida Museum of Natural History to catalog uncataloged ichthyological materials, automate specimen data, and make the ichthyological specimen data accessible electronically over NBII.

#### INFORMATION RESOURCE DIRECTORIES

NBS is currently supporting development of four comprehensive information resource directories that are pertinent to the systematics community. NBS and the Association of Systematics Collections are cooperating in the development of two major directories of taxonomic resource information for North America. One is a directory of taxonomic experts and their respective areas of expertise, the other is a directory of research systematics collections, including information on the electronic accessibility of each collection's specimen data. Both of these directories will be served on NBII and will include direct links to any data or information products that are already available on the Internet.

NBS and The Nature Conservancy (TNC) are cooperating on development of a Natural Heritage Network "node" on the NBII, which will include information on each of the 50 state natural heritage programs in the United States, plus national-level data summaries and information products from TNC, available on-line. NBS and the Fish and Wildlife Information Exchange are cooperating on development of a directory of state-level biodiversity databases, which will also be accessible over NBII.

Each of these directories will obviously provide an accessible source of information on systematics resources (e.g., systematics expertise, systematics collections, state-level biodiversity databases) for people to use. More importantly, each will also provide a baseline view of the current situation in these areas (i.e., what information resources are available to what extent in what areas) and lead to

the identification of additional measures needed to achieve the overall goal of a distributed systematics information infrastructure. For example, they can highlight major taxonomic groups for which significant systematics data are not yet electronically accessible, or new types of tools or protocols needed to support a distributed systematics data network. This information will be used to guide future NBII priorities in the systematics arena. Under a Memorandum of Understanding with the National Science Foundation, NBS hopes to pursue joint efforts to support new projects that focus on these priority areas.

#### DEVELOPING PARTNERSHIPS

As noted above, effective partnerships are key to successful development of the NBII and its systematics component. Partner agencies and organizations maintain most of the significant systematics data; partnerships are needed to jointly develop new software tools and new data and metadata standards; and partnerships are needed to provide the funding resources to support more strategic investments in this area.

One recent example of an exciting partnership in this area is a proposed pilot project among NBS, the Smithsonian's National Museum of Natural History, and the Mexican Commission for the Understanding and Use of Biodiversity (CONABIO). The objectives are to demonstrate and test the feasibility of automating and georeferencing data on selected specimens in the Smithsonian collections that were collected in Mexico, as the basis of developing a distributed "international electronic catalog of Mexican specimens." This project will be done as a pilot to test the methodology and approach and evaluate costs and benefits to the participating agencies. If successful and cost-effective, the approach could be used on other portions of Smithsonian natural history collections and/or in other institutions with Mexican specimens.

#### CONCLUSIONS

The White House Interagency Subcommittee on Biodiversity and Ecosystem Dynamics has identified systematics as a research priority that is fundamental to ecosystem management and biodiversity conservation. This primary need identified by the Subcommittee requires improvements in the organization of, and access to, standardized systematics information on nomenclature, classification, and collections. Working with many partners, NBS is striving to make the valuable data and information embodied in the systematics community, in-



cluding the Nation's natural history collections, more accessible to land use decision makers, resource managers, researchers, students, and the public in general. Through championing the National Biological Information Infrastructure concept, NBS is providing leadership to those collecting and using biological data and information in the

pursuit of informed decision making about the wise use and conservation of the Nation's biological heritage.

#### Literature Cited

National Research Council. 1993. A Biological Survey for the Nation. National Academy Press, Washington, D.C.